Functional analysis of the integrative plasmid PMEA300 of the actinomycete Amycolatopsis Methanolica

Vrijbloed, Jan Willem

The actinomycete *Amycolatopsis methanolica*, subject of our studies on regulation of aromatic amino acid biosynthesis, contains a 13.3 kilobase (kb) plasmid (pMEA300). This thesis reports the complete nucleotide sequence of pMEA300, a detailed analysis of the pMEA300-encoded functions, and the use of a pMEA300-derived *E. coli* shuttle vector for gene cloning in *A. methanolica*.

Site-specific integration of pMEA300

Southern blot analysis revealed that the plasmid pMEA300 is present not only in the free state as ccc DNA but also integrated at a unique location in the genome of *A. methanolica*. The restriction map of the integrated pMEA300 copy was identical to that of the free pMEA300 copy. Unlike the situation for pMEA100 in *A. mediterranei* (Zhu *et al.,* 1990), rearrangements and amplifications of pMEA300 DNA fragments were never detected in *A. methanolica* wild-type.

Integration of pMEA300 in the genome of *A. methanolica* is not at random; recombination occurs between a specific site on the plasmid (*attP)* and a specific site (*attB*) in the chromosome of *A. methanolica*, resulting in formation of the left and right attachment sites of the integrated plasmid (*attL* and *attR*, respectively). All information required for site-specific integration of pMEA300 is carried on a 2.1 kb pMEA300 DNA fragment. The nucleotide sequence of this fragment revealed the *attP* and two open reading frames (ORFs), *int* and *xis*. Based on similarities with known *att* sites, the *attP* of pMEA300 was estimated to consist of 45 bp. Sequencing of the chromosomally located *attR*, *attL* and *attB* sites indicated that the *att* sites in fact encompass 43 bp and that the attL and attB are contained within a putative isoleucine tRNA gene with an ATC anticodon triplet (J.W. Vrijbloed *et al*., unpublished) (Fig. 1). Analysis of deletion derivatives of pMEA300 showed that the two putative proteins Int and Xis are essential for site-specific integration and excision of pMEA300. *Xis* encodes a putative polypeptide of 94 amino acids with a M_r of 9,954; the deduced amino acid sequence of Xis aligns well with other known excisionase proteins. *Int* would encode a protein of 456 amino acids with a M_r of 51,134, showing a high degree of similarity to known recombinases of the integrase family. The data indicate that the pMEA300-encoded sequences for site-specific integration and excision are quite similar to those of other actinomycete integrating plasmids.

Why do pMEA300 and other actinomycete plasmids integrate in the genome, and why does integration occur in tRNA genes? It seems reasonable to speculate that integration in the genome provides these plasmids with additional structural and segregational stability. Less obvious is the reason why pMEA300 and related plasmids integrate in tRNA genes. Clearly, random integration in the genome of newly infected

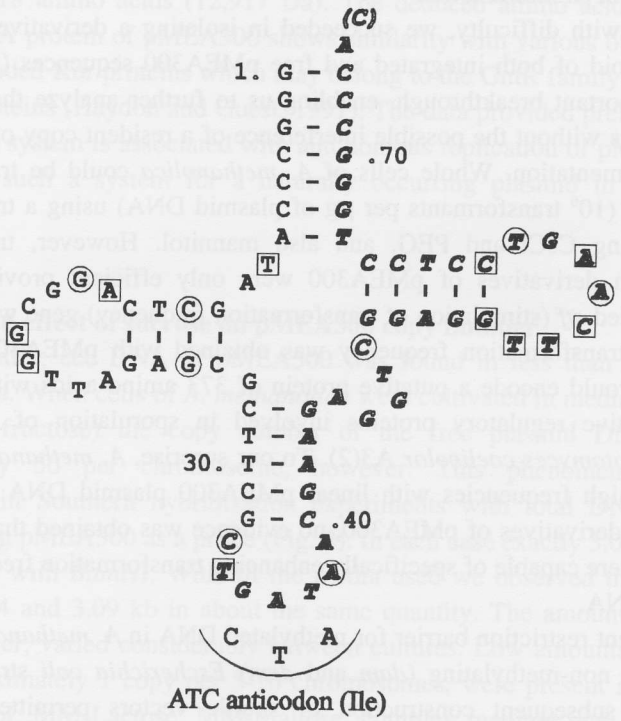

FIG. 1. Cloverleaf representation of the putative Ile-tRNA gene overlapping the pMEA300 *attB* site in *A. methanolica*. Nucleotides shown in italics are conserved in the *attB*, *attP*, *attL* and *attR* sites. Nucleotides in squares (circles) are conserved (semiconserved) positions in tRNA genes.

host cells would in many cases be lethal because essential genes become inactivated. It appears unlikely that the typical tRNA structures are required for the processes of integration and excision catalyzed by Int and/or Xis; integration in non-tRNA like sequences has been reported as well (Brown *et al.,* 1988). Bacterial tRNA sequences are highly conserved and this may provide the true reason for integration of these actinomycete plasmids in tRNA genes. In this way these plasmids may be able to establish themselves, and to be maintained more easily, in cells of different species, which would confer an additional advantage to these plasmids. Experimentally this has been demonstrated by the development of integrative vectors based on the *Streptomyces ambofaciens* plasmid pSAM2 for other *Streptomyces,* but also *Frankia* and *Mycobacterium* species (Alegre *et al.,* 1994; Mazodier *et al.,* 1990).

Transformation of *A. methanolica*

Although with difficulty, we succeeded in isolating a derivative strain of *A. methanolica* devoid of both integrated and free pMEA300 sequences (strain WV1). This was an important breakthrough, enabling us to further analyze the pMEA300-encoded functions without the possible interference of a resident copy of the plasmid by *trans* complementation. Whole cells of *A. methanolica* could be transformed at high frequencies (10^6 transformants per µg of plasmid DNA) using a transformation mixture containing CsCl and PEG, and also mannitol. However, transformation experiments with derivatives of pMEA300 were only efficient provided that the pMEA300-encoded *stf* (stimulation of transformation frequency) gene was present. A 100-fold lower transformation frequency was obtained with pMEA300-derivatives lacking *stf*. *stf* would encode a putative protein of 373 amino acids with M_r 40,201, resembling putative regulatory proteins involved in sporulation of *Streptomyces griseus* and *Streptomyces coelicolor* A3(2). To our surprise, *A. methanolica* could be transformed at high frequencies with linear pMEA300 plasmid DNA. When using various deletion derivatives of pMEA300, no evidence was obtained that pMEA300-encoded genes were capable of specifically enhancing transformation frequencies with linear plasmid DNA.

An efficient restriction barrier for methylated DNA in *A. methanolica* could be avoided using a non-methylating (*dam* and *dcm*) *Escherichia coli* strain for DNA isolations. The subsequent construction of shuttle vectors permitted a detailed functional analysis of pMEA300. Fragments of pMEA300 were cloned in an *E. coli* plasmid and transformed to strain WV1. In this way, regions on pMEA300 encoding functions for site-specific chromosomal integration, autonomous replication, conjugational transfer, stimulation of transformation frequency, high mutation frequency, and regulation were identified.

Autonomous replication of pMEA300

The minimal replicon of pMEA300 was shown to consist of two unlinked DNA fragments of 2.6 kb and 0.8 kb. Sequence analysis of the larger fragment revealed four ORFs. Two of these, *orf36* and *orf49* are very small and a direct role for them in replication is unlikely. *orfA* and *orfB* encode putative proteins of 170 amino acids (18,373 Da) and 416 amino acids (45,260 Da), respectively. No clear similarities were found between the deduced amino acid sequences of OrfA and OrfB and replication proteins of other *Streptomyces* plasmids. The pMEA300 proteins thus may represent unfamiliar types. The replication mechanism of pMEA300 remains to be elucidated. The 0.8-kb DNA fragment contained a single complete ORF (*korA*), encoding a

protein of 118 amino acids (12,917 Da). The deduced amino acid sequence of the putative KorA protein of pMEA300 shows similarity with various other *Streptomyces* plasmid-encoded Kor proteins which may belong to the GntR family of transcriptional repressor proteins (Haydon and Guest, 1991). The data provided preliminary evidence that a *kil-kor* system is associated with autonomous replication of pMEA300, the first example of such a system for a naturally occurring plasmid in a Gram-positive bacterium.

Amplification effect of sucrose on pMEA300 copy number

Generally, ccc DNA of pMEA300 was found in less than 1 copy per 5-10 chromosomes. When cells of *A. methanolica* were cultivated in media with autoclaved sucrose (or fructose) the copy number of the free plasmid DNA increased to approximately 50 per chromosome, however. This phenomenon was further investigated in Southern hybridization experiments with total DNA from various cultures, using pMEA300 as a probe (Fig. 2). In each case exactly 3.0 µg of total DNA was digested with *BamH*I. With all the media used we observed the two additional bands of 23.4 and 3.09 kb in about the same quantity. The amount of free plasmid DNA, however, varied considerably between cultures. Low amounts of free plasmid DNA, approximately 1 copy per 5-10 chromosomes, were present in cells grown in MM-Methanol, MM-Lactose, MM-Maltose, complex medium and YEME-Mannitol media. At least 500-1000 times more free pMEA300 was present in cells grown in YEME, MM-Sucrose, MM-Fructose media (Fig. 2), confirming the results from plasmid isolation studies. Large amounts of free plasmid DNA can thus be obtained from cells grown in media with sucrose and, to a lesser extent, fructose. When using filter-sterilized sucrose the pMEA300 copy number remained at its normal low level of less than 1 free copy per chromosome. The high copy number of pMEA300 thus depends specifically on the use of autoclaved sucrose (J.W. Vrijbloed *et al.*, unpublished). High copy number derivatives of pSAM2, coexisting with an integrated copy, have been reported following UV irradiation (Pernodet *et al.,* 1984; Sezonov *et al.,* 1995). However, pMEA300 appears to be the first example of a plasmid that can be amplified using simple variations in medium composition. The amplification is probably induced by specific degradation products of sucrose (and fructose) that accumulate during autoclaving. A general stress response due to the accumulation of unknown toxic compounds during autoclaving cannot be excluded.

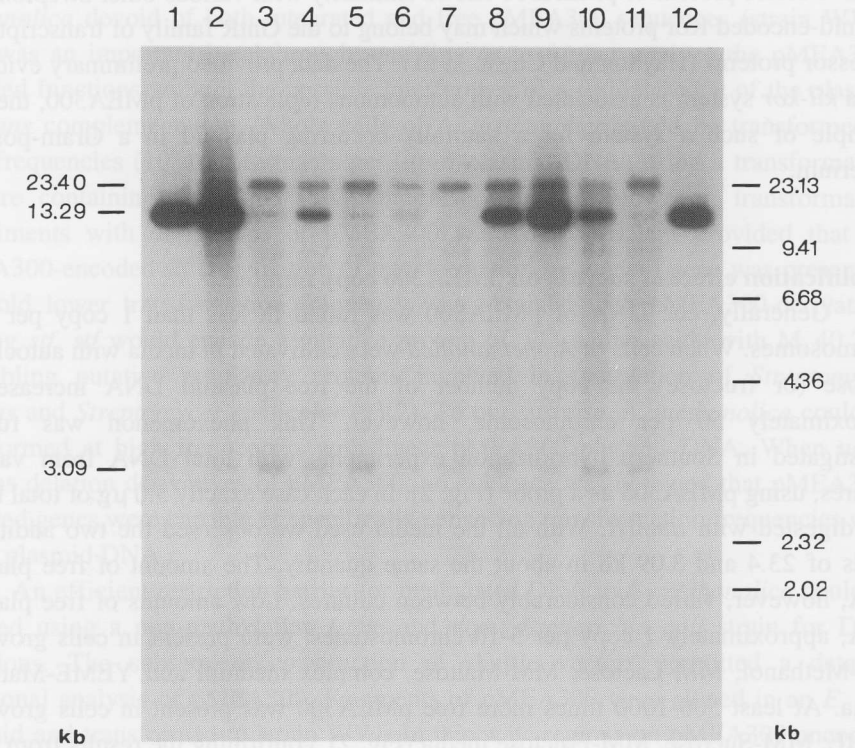

FIG. 2. Southern hybridization of ^{32}P-labelled pMEA300 DNA with *Bam*HI digests of total DNA of *A. methanolica*. Total DNA samples were isolated from cells grown on various media; lane 2, YEME; lane 3, YEME (sucrose omitted) supplemented with PEG; lane 4, YEME (sucrose omitted) supplemented with Mannitol; lane 5, CM; lane 6, Minimal medium (MM) supplemented with Maltose; lane 7, MM supplemented with Lactose; lane 8, MM supplemented with Sucrose; lane 9, MM supplemented with Fructose; lane 10, MM supplemented with Glucose; lane 11, MM supplemented with Methanol. Linearized pMEA300 DNA was used as a control (lanes 1 and 12).

Conjugative transfer of pMEA300

In *A. methanolica* pMEA300 is capable of conjugational transfer from plasmid-carrying cells to plasmid-lacking cells. The transfer of pMEA300 results in the formation of pocks. Pocks are zones of growth inhibition that become apparent when plasmid-carrying donor cells develop in a confluent lawn of plasmid-lacking recipient cells. A 6.2-kb DNA region on pMEA300 specified the functions of conjugational

transfer and pock formation. Sequencing of the 6.2-kb DNA region revealed ten ORFs. This is the first sequence of the transfer region of a plasmid from a non-streptomycete actinomycete. A detailed mutational analysis of this region showed that at least four individual proteins, TraG (9,488 Da), TraH (12,586 Da), TraI (40,468 Da) and TraJ (81,109 Da), are required for efficient transfer of pMEA300. Their disruption resulted in a clear reduction in the transfer frequencies, ranging from 5.2 x 10^1 fold (TraG) to 2.3 x 10^6 fold (TraJ), and in reduced pock sizes. At least two putative proteins, TraA (10,698 Da) and TraB (31,442 Da), are responsible for pock formation specifically. No clear similarities were found between the deduced sequences of the ten putative Tra-proteins of pMEA300 and those of self-transmissible, (non) integrating *Streptomyces* plasmids studied. All the Tra-proteins of pMEA300 thus may represent unfamiliar types. However, similar to the situation in pSAM2 and other non-integrating *Streptomyces* plasmids, an ATP binding box and a putative DEAD box were found in TraJ. This suggests that TraJ functions as an DNA helicase (Linder *et al.,* 1989) and that the mechanism of transfer of pMEA300 is quite similar to other *Streptomyces* plasmids. The detection of one to three membrane-spanning regions in the putative *traE*, *traG-traJ* gene products of pMEA300 suggested that these proteins are membrane-associated. A detailed analysis of the mechanism of pMEA300 transfer and the function of each of these proteins remains to be carried out.

Many pock forming plasmids are also fertility plasmids, capable of mobilizing chromosomal DNA markers (Hopwood and Kieser, 1993). In matings with *A. methanolica* strain WV1 derived (pMEA300-free) strains mobilization of chromosomal markers still occurred (Euverink *et al.,* 1996). This could mean that pMEA300 is not, or not the only DNA element, involved in chromosome mobilization in *A. methanolica.* Isolation of a strain completely deficient in fertility followed by reintroduction of pMEA300 may serve to elucidate whether pMEA300 can in fact mobilize chromosomal markers.

A pMEA300-encoded system for high mutation frequency

A very interesting feature of pMEA300 was discovered when wild-type and strain WV1 were incubated on glucose minimal medium agar plates containing 1 mg/ml para-fluoro-phenylalanine (pFPhe). After incubation at 37°C, numerous pFPhe-resistant mutants of wild-type had appeared on these agar plates. However, the pMEA300-free strain WV1 showed an almost 300-fold reduced frequency of spontaneous mutagenesis. The high mutation frequency could be restored by re-introducing pMEA300 in strain WV1. We concluded that pMEA300 encodes a system for high frequency of spontaneous mutagenesis. Subsequent deletion analysis of

pMEA300 revealed that *mut*, encoding a putative protein of 179 amino acids with a molecular weight of 20,259, was responsible for this high mutation frequency. Database searches revealed that *mut* contains the highly conserved sequence motif defining the MutT family of proteins. Autonomous replication of pMEA300 is essential for effective expression of the mutator phenotype: the presence of a single integrated copy of *mut* fails to support a high mutation rate.

The *E. coli mutT* gene encodes a DNA repair system. Its inactivation results in a drastic increase in the frequency of spontaneous mutations. This is in clear contrast to the situation observed for the pMEA300-encoded Mut system in *A. methanolica*. It thus appears unlikely that the latter system is involved in DNA repair. Rather, the pMEA300 mutator system may play a role in genotypic adaptation processes of the host, increasing its chances of survival under stress conditions in the natural environment. Mesocosm experiments with *A. methanolica* wild-type and strain WV1 are currently carried out (collaboration with E.M. Wellington, University of Warwick), to try and obtain more evidence for the importance of such a mutator system in the natural environment.

The mechanism of mutagenesis via the pMEA300 Mut system remains to be elucidated. Cloning and sequencing of genes coding for deregulated (feedback inhibition insensitive) prephenate dehydratase proteins from several independently isolated pFPhe resistant mutant strains of wild-type is well under way now (H.J. Kloosterman *et al.*, unpublished). It it expected that such nucleotide sequence data will provide further insights about the mechanism of the pMEA300-encoded mutator system.

Regulation of pMEA300-encoded functions

Several observations led to the notion that the various functions encoded by pMEA300 are under strict regulatory control. (1) Autonomous replication of pMEA300 requires the presence of the *korA* gene (Vrijbloed *et al.,* 1995a) encoding a putative repressor protein of the GntR family; (2) the *stf* gene is needed for efficient transformation of *A. methanolica* cells with pMEA300 (Vrijbloed *et al.,* 1995b), and (3) deletion of *mut* is lethal in case of *stf* containing plasmids, but not when both *mut* and *stf* are deleted (Vrijbloed *et al.,* 1996). To investigate the roles of the *mut*, *stf* and *korA* gene products in controlling the various pMEA300-encoded functions, a series of deletion plasmids based on pWV129 was constructed (Fig. 3). The deletion plasmids were transformed to strain WV1 and the effects of the deletions on replication, transformation frequency, copy number, pock formation and mutation frequency were determined (Fig. 3).

amino acids with a
nutation frequency.
ved sequence motif
n of pMEA300 is
resence of a single

activation results in
is in clear contrast
n *A. methanolica*. It
repair. Rather, the
on processes of the
ons in the natural
pe and strain WV1
ersity of Warwick),
tator system in the

stem remains to be
egulated (feedback
eral independently
der way now (H.J.
sequence data will
0-encoded mutator

ctions encoded by
us replication of
1995a) encoding a
eeded for efficient
et al., 1995b), and
not when both *mut*
of the *mut*, *stf* and
nctions, a series of
deletion plasmids
ns on replication,
on frequency were

Characteristics	Deletion	Replication	Transformation freq.	Copy number CM, YEME	Pocks	Mutation freq
pMEA300	none	+	normal	low, high	normal	high
pWV129	int	+	normal	high, lethal	normal	high
pNAT102	mut	-	n.a.	n.a.	super	n.a.
pWV374	mut	-	n.a.	n.a.	super	n.a.
pWV113	mut, stf	+	reduced	low, high	normal	low
pWV111	mut, stf	+	reduced	low, high	normal	low
pWV375	stf	+	reduced	low, high	normal	high
pWV383	mut, stf, korA	-	n.a.	n.a.	super	n.a.

FIG. 3. Restriction map of pWV129 (Vrijbloed *et al.*, 1995a) and functional analysis of pWV129 derived plasmids. Only restriction sites relevant for the construction of the plasmids are shown. Numbers indicate the positions (in bp) of the restriction sites from the unique *Eco*RI site. Arrows indicate open reading frames on pWV129 (Vrijbloed *et al.*, 1994; Vrijbloed *et al.*, 1995a,b; Vrijbloed *et al.*, 1996); the dotted arrow and box, vector pHSS6-derived kanamycin (*km*) resistance gene and *colE1* origin. Arrowheads indicate restriction sites that have been disrupted with KLENOW. Bars indicate the various deletions in pWV129. The results of attempts to transform strain WV1 with these plasmids, selecting for pock formation or resistance against kanamycin, are summarized underneath. n.a., not applicable.

Plasmids pWV374 and pWV375 carrying the *stf* gene, but with a deleted or disrupted *mut* gene, form so-called "super" pocks. These pocks are characterized by the absence of a colony of viable cells within the lysis zone, a phenomenon that is probably caused by uncontrolled expression of pMEA300-encoded *kil*-genes, most likely *traA/traB* (Vrijbloed *et al.,* 1995c) and one or more of the replication genes (Vrijbloed *et al.,* 1995a). The additional deletion of *stf* on these plasmids (pWV113, pWV111) restored the ability to form normal pocks, although with a lower transformation frequency. The pMEA300-Mut protein displays homology to proteins capable of hydrolysing ATP/GTP, or even NADH (Frick *et al.,* 1995; Frick and Bessman, 1995). It is speculated that Mut inactivates Stf by hydrolysis of a bound ATP (or GTP/NADH). Deletion of *mut* from *stf* containing plasmids (e.g. pWV374) thus may result in a highly active form of the Stf protein. This, and a direct regulatory effect of the Stf protein on transcription from the *traA/traB* promoter, remains to be shown, however.

KorA may be involved in the regulation of the medium composition-dependent amplification of free pMEA300. The amplifying effect of autoclaved sucrose was still observed with a deletion derivative of pWV129 that carried only the minimal replicon (J.W. Vrijbloed *et al.*, unpublished). KorA may be involved in the amplification process as a transcriptional repressor protein. Members of the family of GntR proteins contain an N-terminal DNA binding domain and a C-terminal effector binding domain. As shown in the analysis of gluconate utilization by *Bacillus subtilis*, the effector molecule in the case of GntR is gluconate (Haydon and Guest, 1991). Binding of gluconate to GntR releases the protein from the DNA and this leads to full expression of the gluconate utilization genes. It can be speculated that in case of KorA of pMEA300 the effector molecule is a specific degradation product that accumulated during autoclaving of sucrose. Its interaction with KorA thus would result in a higher expression of the replication (and transfer) genes and subsequently to a higher copy number (and an increased transfer frequency) of pMEA300.

Quite surprisingly, plasmid pWV129 displayed a high copy number in complex medium (Fig. 3; J.W. Vrijbloed *et al.*, unpublished) where a low copy number is observed for pMEA300. Although speculative, this may be explained by a relatively poor expression of *mut* on pWV129 due to the upstream insertion of the *E. coli* plasmid pHSS6. This could result in the formation of more active Stf protein. On YEME medium this results in a lethal phenotype, most likely because of lack of control of the *traA/traB* promoter by KorA and Stf (Fig. 3). A hypothetic model of proteins and factors that may be involved in control of transcription of the *traA/traB* promoter is presented in Fig. 4.

How is transcription of the pMEA300 regulatory genes *korA*, *stf* and *mut* regulated?

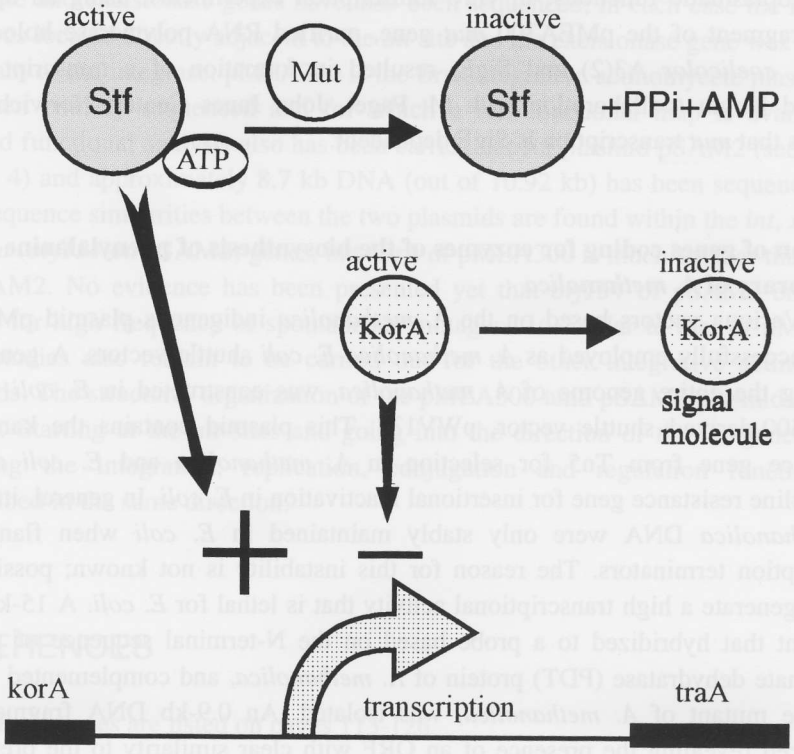

FIG. 4. Schematic model of the regulation of the *traA/traB* promoter at the level of transcription by KorA, Stf and Mut.

Recently we succeeded in overexpressing the KorA protein in *E. coli* and showed that it binds to the *korA-traA* intergenic region (Vrijbloed *et al.,* 1995c). Interestingly, similar inverted repeats of 14 bp containing an *Avr*II restriction site in the central core were identified upstream of *korA*, *traA* and *orfA*. Such a sequence was previously identified as the binding site of ImpA (Shiffman and Cohen, 1993), a homolog of KorA encoded on SLP1. Conceivably, the KorA protein thus not only
plays a role in transcriptional control of the pMEA300 *rep* and *tra* genes, but also controls its own expression as an autoregulator. Presently we do not know whether *korA* and *stf* are cotranscribed, or that *stf* is transcribed seperately.

Upstream of the *mut* gene we discovered a potential binding site for the sigmaE factor (Lonetto *et al.,* 1994). SigE is suggested to be involved in the regulation of

extracytoplasmatic functions. In vitro transcription experiments, using an upstream DNA fragment of the pMEA300 *mut* gene, purified RNA polymerase holoenzyme from *S. coelicolor* A3(2) and SigE, resulted in formation of a transcript of the expected length (collaboration with M. Paget, John Innes Center, Norwich). This suggests that *mut* transcription is SigE dependent.

Isolation of genes coding for enzymes of the biosynthesis of phenylalanine from a gene library of *A. methanolica*

Various vectors based on the *A. methanolica* indigenous plasmid pMEA300 were successfully employed as *A. methanolica-E. coli* shuttle vectors. A gene bank, covering the entire genome of *A. methanolica,* was constructed in *E. coli* using a pMEA300-derived shuttle vector, pWV138. This plasmid contains the kanamycin resistance gene from Tn5 for selection in *A. methanolica* and *E. coli* and the tetracycline resistance gene for insertional inactivation in *E. coli*. In general, inserts of *A. methanolica* DNA were only stably maintained in *E. coli* when flanked by transcription terminators. The reason for this instability is not known; possibly the inserts generate a high transcriptional activity that is lethal for *E. coli.* A 15-kb DNA fragment that hybridized to a probe based on the N-terminal sequence of purified prephenate dehydratase (PDT) protein of *A. methanolica,* and complemented a PDT-negative mutant of *A. methanolica,* was isolated. An 0.9-kb DNA fragment was sequenced revealing the presence of an ORF with clear similarity to the prephenate dehydratase-encoding genes of other organisms. The *A. methanolica pdt* gene encodes a 304 amino acid protein with a molecular weight of 32,288. Subsequent work has also resulted in cloning and characterization of the *A. methanolica* genes encoding the other key regulatory enzymes in the biosynthetic pathway for L-phenylalanine (DAHP synthase and chorismate mutase; J.W. Vrijbloed *et al.*, unpublished; H.J. Kloosterman *et al.*, unpublished). Also further genes involved in methanol (*mno* encoding methanol dehydrogenase; H.J. Hektor *et al.*, unpublished) and glucose (*pfp* and *pfk* encoding PPi-dependent and ATP-dependent phosphofructokinases, *pgm* encoding phosphoglycerate mutase; A.M.C.R. Alves *et al.*, unpublished) have been cloned from *A. methanolica.*

Integrative plasmids in actinomycetes

To date nine different integrative actinomycete plasmids are known; for most of these plasmids the *att* sites have been sequenced (see chapter 1: Table 1). Without exception these plasmids integrate in a putative tRNA. For four of these plasmids the

integrase and excisionase genes have also been sequenced; in each case the integrase gene was located directly adjacent to the *att* site and the excisionase gene was encoded upstream of the integrase. pMEA300 is the first integrative actinomycete plasmid that has been entirely sequenced and for which a full functional map is available. A detailed functional analysis also has been carried out for plasmid pSAM2 (see chapter 1: Fig. 4) and approximately 8.7 kb DNA (out of 10.92 kb) has been sequenced. The only sequence similarities between the two plasmids are found within the *int*, *xis*, *korA* and *mut* (*orf154* on pSAM2) genes, but *korA* of pMEA300 is much smaller than *korSA* of pSAM2. No evidence has been presented yet that *orf154* of pSAM2 encodes a system for high frequency of spontaneous mutagenesis similar to that of pMEA300. Such studies also remain to be carried out for the other integrative actinomycete plasmids. The structural organization of the pMEA300 amd pSAM2 plasmids is quite similar. Starting at the *att* sites and going into the direction of the *int* genes, genes encoding the integration, replication, conjugation and regulation functions are transcribed in the same direction.

REFERENCES

References are listed on pages 113-126.